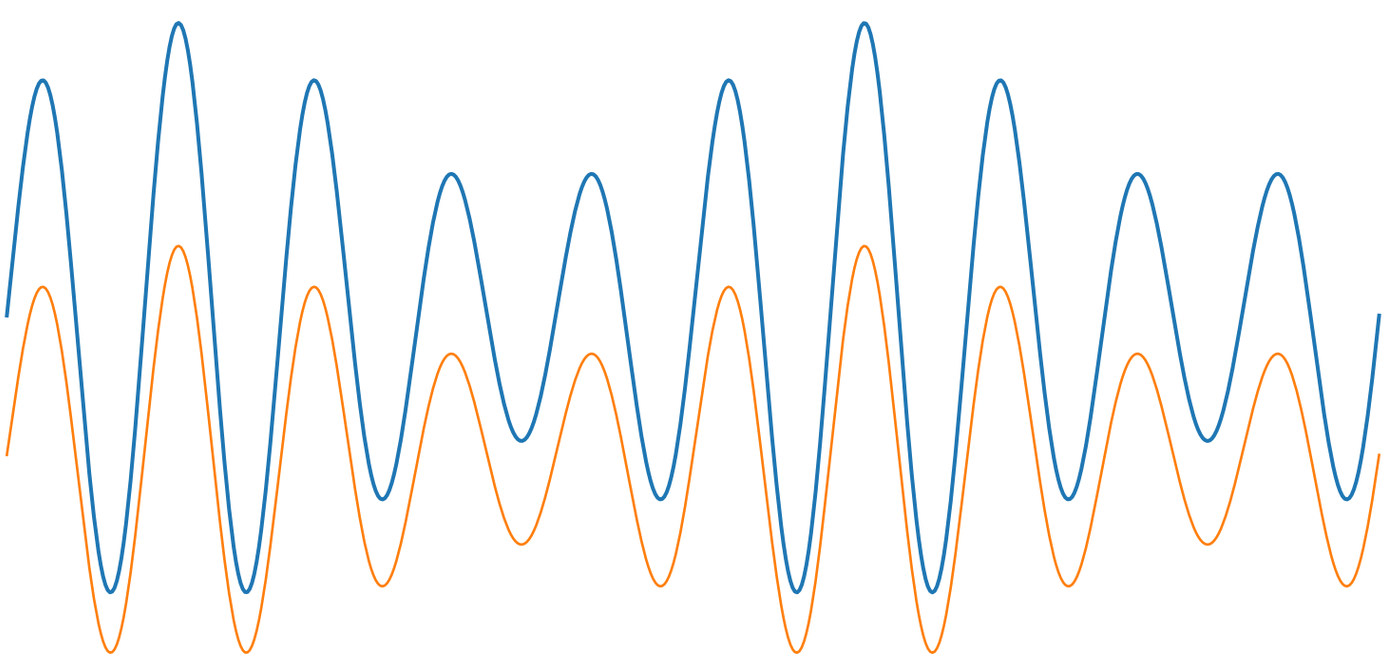
WhisperMic — Origin, Design, and Business Case

Pre-roll Buffer • Ultra-light VAD Markers • Optional Subaudible Silence Flag

A mic-side solution to reduce clipped speech and remove redundant AI analysis of silence

Prepared by: Liam Fahey (with AI co-design assistance)
Date: August 13, 2025
Location: Ireland



Provenance & Ideation Timeline (from chat sessions)

- Aug 7, 2025 — Initial problem framing: clipped first words after unmute/PTT; interest in tagging noises so editors can jump to them in the waveform.
- Aug 7, 2025 — Two-channel idea proposed: (1) binary non-silence flag, (2) optional coarse tags (speech/noise/impact). Name ‘WhisperMic’ adopted.
- Aug 7–8, 2025 — Feasibility thread: mic-side circular pre-roll buffer (100–500 ms), ultra-light VAD (RMS + ZCR/spectral-flux), HID/vendor sideband at ~20–50 Hz.
- Aug 8, 2025 — Audio-only fallback proposed: optional subaudible control tone during confirmed silence to signal ‘do not analyse’; default-off with safety gates.
- Aug 8, 2025 — Per-user cost sketch (time, storage) and first macro model for AI waste on silence. Discussion of G HUB controls and DAW/NLE marker ingestion.
- Aug 9–10, 2025 — Refinements: privacy (on-device), sample-rate guardrails (disable tone <48 kHz), marker taxonomy kept intentionally simple.
- Aug 12–13, 2025 — Business case tightened: 1.5–3.0B/year global waste estimate; request to prepare a PDF with diagrams and support-letter for Logitech.

Key attributions (AI-led suggestions):

– Pre-roll buffer on device to prevent clipped onset.

– Low-rate HID/vendor markers for non-silence + coarse events.

– Optional subaudible control tone as a metadata-free fallback.

– G HUB control mockups; DAW/NLE marker ingestion path.

– Conservative macro cost model (1.5–3.0B/yr) and savings math.

Executive Summary

- Problem: host apps and cloud AIs waste cycles on dead air; users suffer clipped first syllables after unmute/PTT.
- Fix: Mic-side pre-roll + optional low-rate markers; fallback subaudible silence flag for audio-only paths.
- Impact: Better UX, lower storage/egress, and less redundant AI analysis before it leaves the device.
- Scale: Reducing silence entering AI pipelines by only 10–20% yields 150*M*–600M/year in global savings.

Core Feature: Pre-roll Buffer

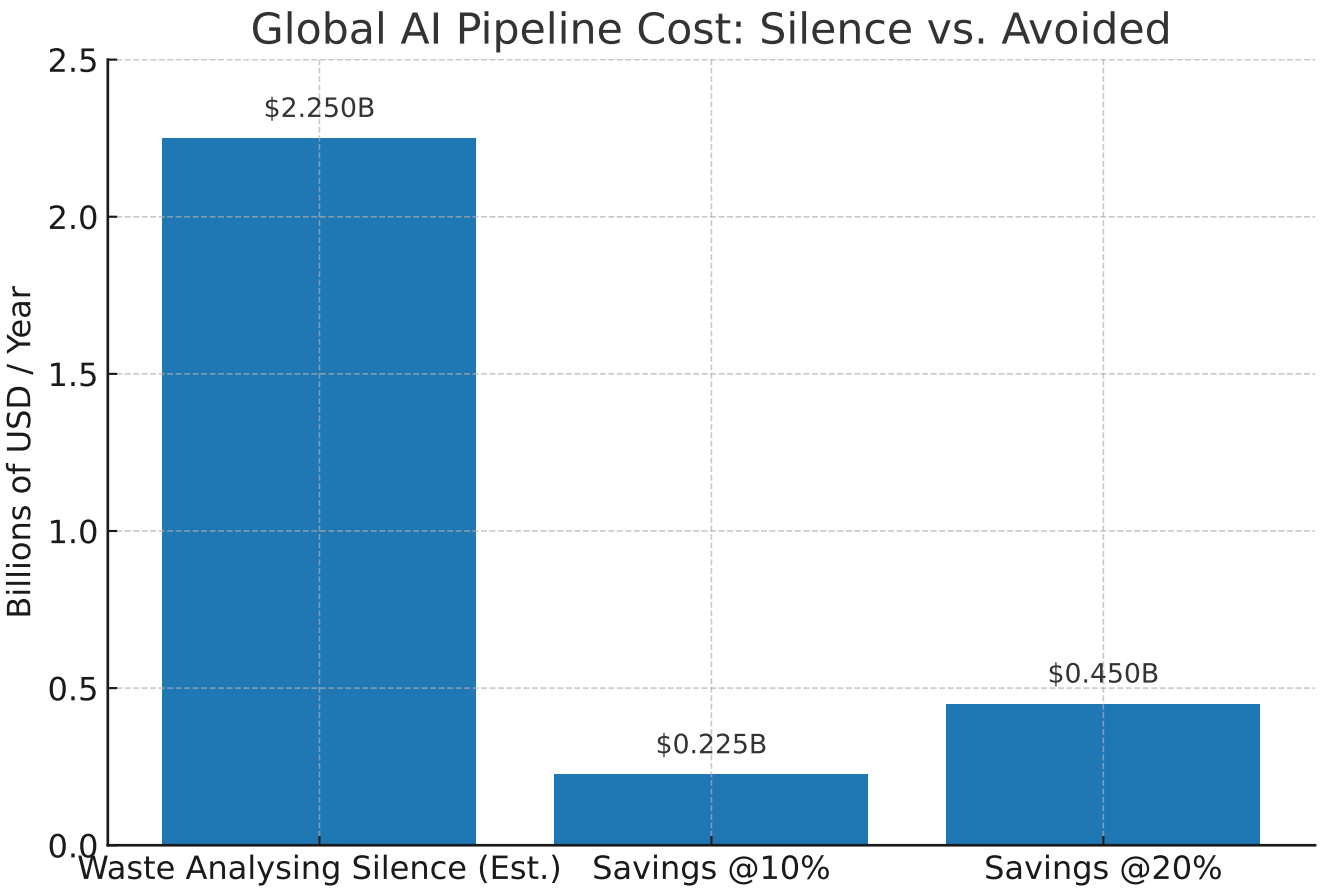
- Circular PCM buffer; prepend on unmute/PTT/VOX so the first syllable is captured.
- G HUB: on/off + pre-roll length slider (100–500 ms).

Optional: Whisper+ Markers

- HID/vendor endpoint (20–50 Hz) flags non-silence and coarse events (speech/noise/impact).
- Live apps gate smarter; DAWs auto-mark timeline. Content-free booleans/labels.

Optional: Subaudible Silence Flag

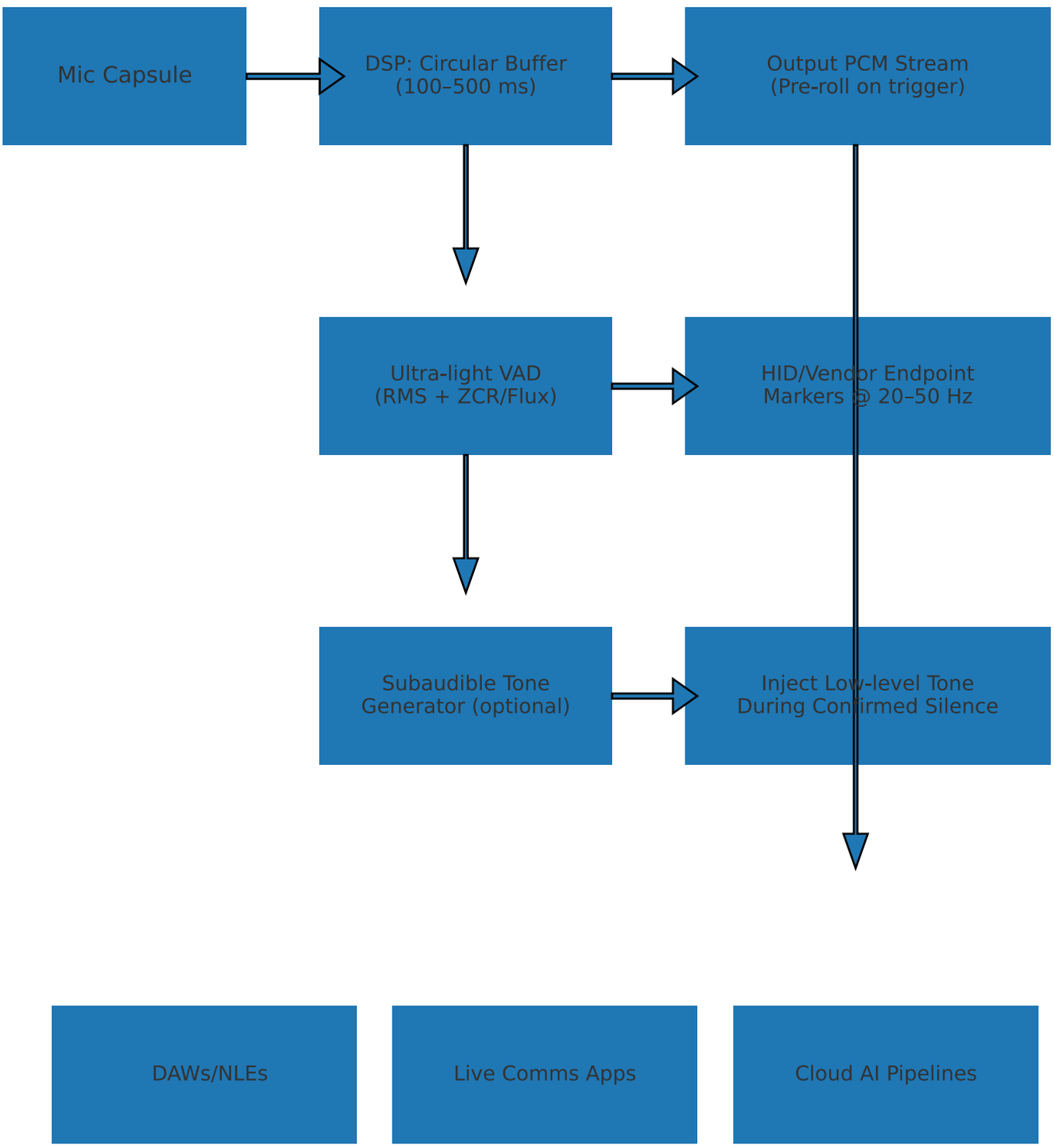
- Very low-level, windowed control tone during confirmed silence (default-off; amplitude-limited).
- Auto-disable under incompatible sample rates or safety constraints.



Assumptions (Conservative)

- Total audio hours analysed by AI per year: ~10.5B hours.
- Dead air proportion still processed: 12–18%.
- All-in compute per analysed hour (VAD/ASR/diarisation at scale): 1.2–1.6.
- Estimated annual waste: 1.5–3.0B (midpoint shown).
- Mic-side changes reduce silence before it leaves the device, cutting redundant AI analysis.

System Architecture (Mic → Apps/Cloud)

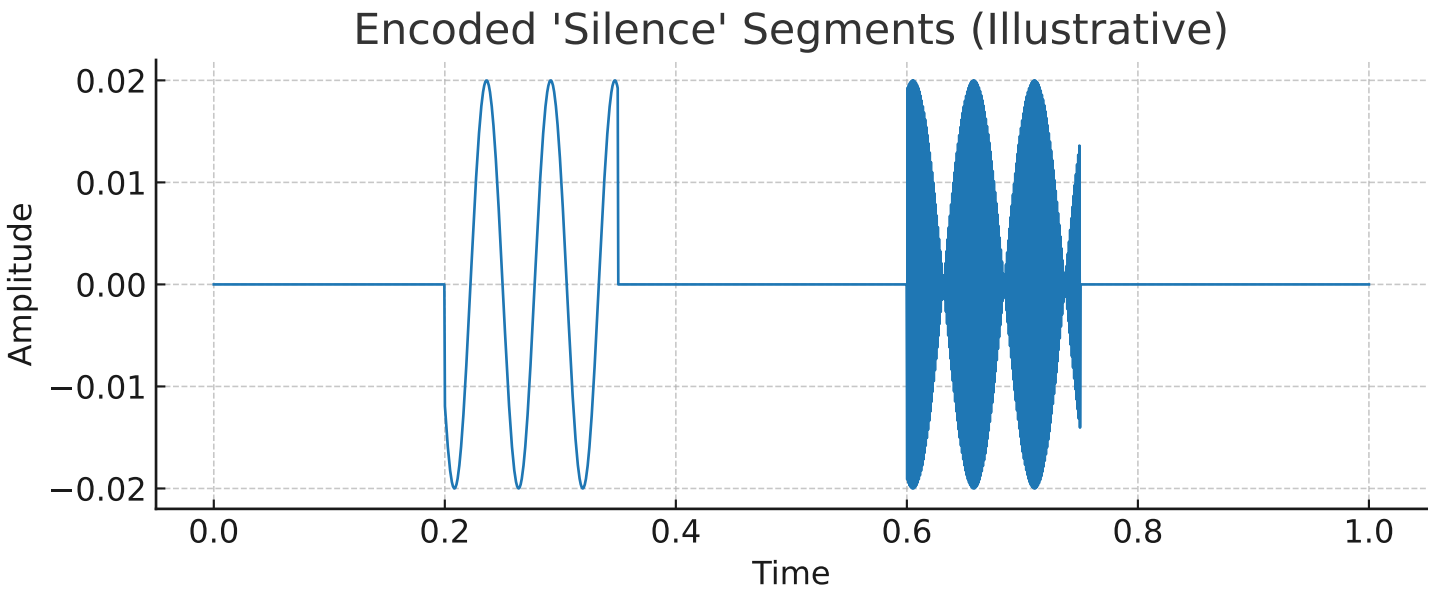


Optional Subaudible Silence Flag (Audio-only Environments)

- Purpose: For pipelines that cannot ingest metadata, insert a control tone during confirmed silence to signal 'do not analyse'.
- Variants (amplitude-limited, windowed): LF ~17–20 Hz (may be high-pass filtered) or HF ~19 kHz (potential audibility on some gear/users).
- Safeguards: Default-off; target level ~−50 dBFS; smooth fade in/out; disable if sample rate < 48 kHz; hearing-safe constraints enforced.
- Preference: Use HID/Vendor markers whenever possible; tone is a fallback only.

LF control tone (~18 Hz)

HF control tone (~19 kHz)



G HUB Controls (Concept Mock)

Pre-roll Buffer

Enable: [ ] Length: [250 ms ▼] Range: 100–500 ms

Whisper+ Markers

Enable: [ ] Rate: [25 Hz ▼] Tags: [Non-silence] [Impact] [Broadband Noise]

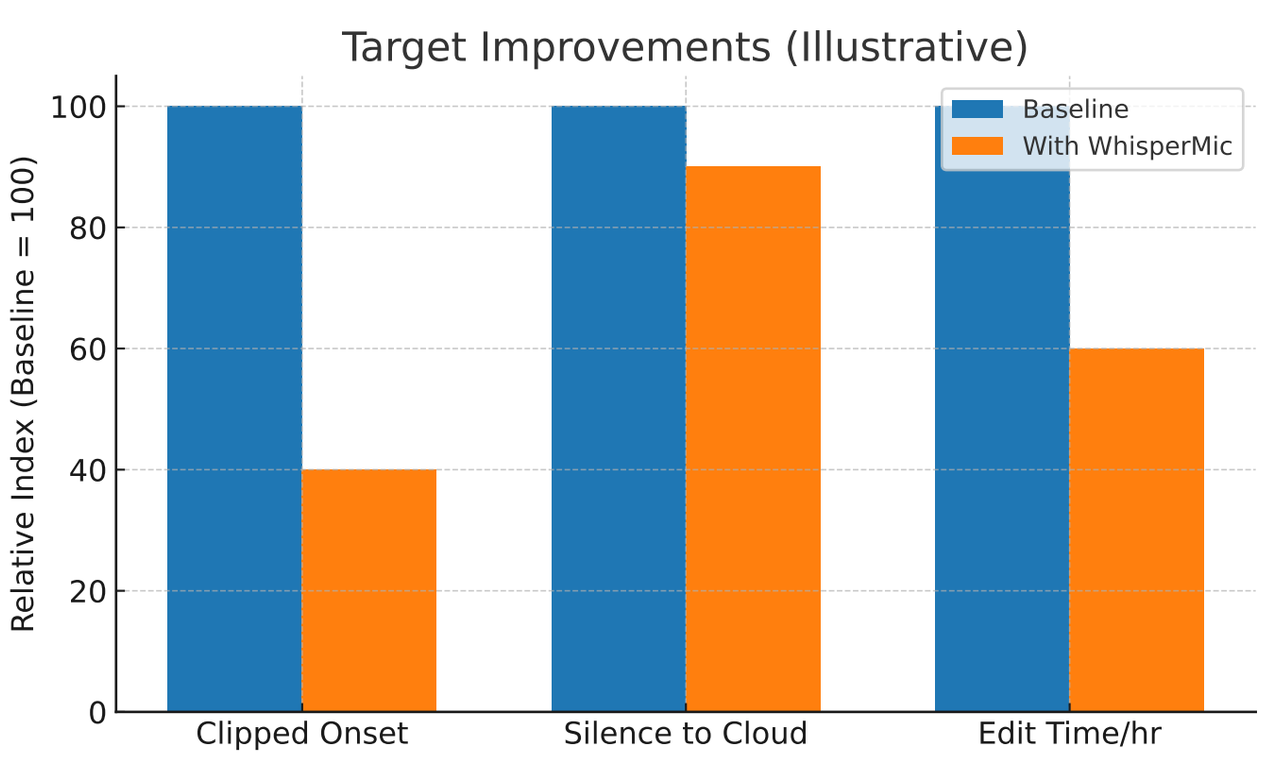
Subaudible Silence Flag (Fallback)

Enable: [ ] Mode: [LF 17–20 Hz ▼ / HF 19 kHz ▼] Level: [−50 dBFS ▼]

Auto-disable if sample rate < 48 kHz or hearing-safe constraints not met.

Measurement & Validation Plan

- A/B test in G HUB beta: baseline vs. pre-roll; baseline vs. markers; combined effects.
- Metrics: clipped-onset rate (% of utterances starting above −24 dBFS within first 50 ms), AI billable minutes, VAD non-silence ratio, human edit minutes/hour.
- Data capture: on-device telemetry (aggregated/anonymous) + opt-in user surveys for edit time and perceived intelligibility.
- Success thresholds (illustrative): ≥60% reduction in clipped-onset, ≥10% reduction in silence shipped to cloud, ≥5 min/hour edit time saved.



Risks & Mitigations

- Audible artefacts: Windowed pre-roll splice; safety-checked tone levels; auto-disable under incompatible conditions.
- False VAD positives: Conservative thresholds; hysteresis; allow app override; keep taxonomy minimal.
- Chain compatibility: Markers prefer HID; tone is fallback; clear docs for integrators; sample-rate guards.
- User confusion: Off-by-default; short tooltips; presets; easy rollback.
- Privacy concerns: All on-device; markers are content-free booleans/labels; no cloud dependency.

The Ask

- Feasibility: Can pre-roll be added to existing USB mics via firmware/driver, or best for a new SKU?
- Beta: If feasible, provide a G HUB beta toggle for pre-roll and markers to validate metrics at scale.
- Hardware path: If firmware-limited, consider a creator-focused SKU including these features.
- Integration: Share preferred SDK/plugin path for ingesting markers in DAWs/comms apps.

Contact

Liam Fahey • Ireland • [Email] • [Optional phone]
Attachment: WhisperMic_Proposal_v2_Liam_Fahey.pdf